A Combination between Textual and Visual Modalities for Knowledge Extraction of Movie Documents

Manel Fourati, Anis Jedidi and Faiez Gargouri
MIR@CL Laboratory, University of Sfax, Sfax, Tunisia

Keywords: Description, Segmentation, Movie Document, Textual, Visual Modalities, Knowledge, LDA.

Abstract: In view of the proliferation of audiovisual documents and the indexing limits mentioned in the literature, the progress of a new solution requires a better description extracted from the content. In this paper, we propose an approach to improve the description of the cinematic audiovisual documents. However, this consists not only in extracting the knowledge meaning conveyed in the content but also combining textual and visual modalities. In fact, the semiotic desription represents important information from the content. We propose in this paper an approach based on the use of pre and post production film documents. Consequently, we concentrate efforts to extract some descriptions about the use not only of the probalistic Latent Dirichlet Allocation (LDA) model but also of the semantic ontology LSCOM. Finally, a process of identifying a description is highlighted. In fact, the experimental results confirmed the importance of the performance of our approach through the comparison of our result with a human jugment and a semi-automatic method by using the MovieLens dataset.

1 INTRODUCTION

Audiovisual documents provide a wide range of content descriptions across different descriptors of different types of media. Therefore, the extraction of these descriptions has received increased attention by several researchers (Tang et al., 2019) (Sanchez-Nielsen et al., 2019). Besides, the process requires some techniques for the extraction of information from the content. All this has an impact on filmic documents which are rich in content and therefore can be a source of complete description. Although, various techniques have been proposed for this process, semantic descriptions are still lacking. In fact, the descriptions of audiovisual documents must be enriched from the content while exploiting the knowledge provided in the document.

Knowing that our objective in this paper is to extract semiotic descriptions from the content, we propose a solution for the extraction of descriptions by applying a multimodal approach based on an automatic process. Therefore, our focus is on the description of filmic document. A film product passes through three phases, namely, a pre-production phase, a production phase and a post-production phase. In our work, we take advantage of these phases to exploit the following documents about the extraction of semiotic descriptions:

- The textual documents used in the filmic preproduction phase, namely, the script.
- The audiovisual stream in its phase of filmic postproduction through the text superposed on the images.
- The cinematographic structure of the audiovisual document in the pre and post-production phase.

Based on these documents, the proposed approach helps extract hidden knowledge from the content. Therefore, we propose the use of two types of analyses, the first is statistical while the second is semantic. On the statistical level, we focus on the use of the Latent Dirichlet Allocation (LDA) probabilistic model while on the semantic level, we opt for the LSCOM ontology. Thereafter, we propose a process to identify the pertinent description related to the movie based on the pertinence and the weighed measure.

In what follows, we will present our proposed approach for the extraction of semiotic descriptions. In section 2, we will present a background of the semiotic description used in this paper and an overview of the methods proposed in the literature, which are related to the description and identification of topics and themes in audiovisual documents. Then, in section 3, we will discuss our solution for the automatic extraction and identification of the thematic description. Subsequently, section 4 will be devoted to the

Fourati, M., Jedidi, A. and Gargouri, F.
A Combination between Textual and Visual Modalities for Knowledge Extraction of Movie Documents.
DOI: 10.5220/0008494902030214
In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, pages 203-214
ISBN: 978-989-758-382-7

already carried out presentation of the various experiments.

2 RELATED STUDIES

2.1 Background

The semiotic has received great attention in the last decades by many researchers in several disciplines, such as, narrative, image and audiovisual semiotics. In our research, we are interested mainly in the audiovisual semiotics, which is defined as a process in which the meanings related to the content of audiovisual documents are taken into account (Martin, 2005). In this respect, Peter Stockinger (Stockinger, 2003), (Stockinger, 2011) and (Stockinger, 2013)), was interested, in his researchs studies, in the audiovisual semiotics of movie documents. Indeed, he proved that the audiovisual semiotics helps approach the semantics of the content and the history it presents. In fact, it is in this respect that our research was conducted. It defines different thematic descriptions, such as, such as dominant themes, discourse themes, taxeme and specified themes.

- Dominant Theme: It represents the pertinent themes in the audiovisual segment, which is considered as the topic of this segment.
  - Discourse Theme: It represents the different theme treated in the audiovisual segment. other words, the theme is used to identify the semantic space of an audiovisual segment.
  - Taxeme: It represents the most relevant theme of discourse in the topic or dominant theme.
  - Specified Theme: It allows the identification of each term related to the theme of the discourse identified.

2.2 Description of Topics in Audiovisual Documents

The goal behind a better description that reflects the content of the document, is undoubtedly, a reliable modeling of the content. In fact, the need for using it is to obtain the required knowledge that takes into account the growth of the interest of the description of the audio-visual documents. The objective is to extract descriptions of the audiovisual content of a ”film” nature.

In addition to the keywords extracted from the content, the topics and themes represent an important source of knowledge extracted from the film. Indeed, these themes are classified among the most natural descriptions inspired by a user. Several efforts have been made to extract this type of description from the content of the film documents.

wWe focus in our state of the art on the presentation of the state of the art studies related to the filmic and textual document. Then, in the framework of the thematic description of the filmic documents, a rather primitive solution has already been used in the literature. It consists mainly in extracting subjects through low-level features.

Recently, the authors in (Bougiatiotis and Giannakopoulos, 2016) have proposed a method of modeling subjects of films based on low level characteristics. Such a method aims at combining different modalities, namely, the subtitles of the films, the audio characteristics and the generic metadata. In order to carry out this fusion process, the authors suggested defining a characteristic vector of each modality. For the first modality, the probabilistic model ”LDA” is applied to the bag of words (BoW) resulting from a pre-processing step on the subtitles. The second modality focuses on on the classification of audio segments to audio events and musical genres. Finally, the third modality makes it possible to select a generic description set from Imdb, namely, genre, directors, etc. The basic goal of topic modeling is, roughly, to measure the similarity between the films. Each film is presented as a label vector to create a ”Ground Truth” similarity matrix between the films. The proposed method is tested on a basis of 160 films selected from Imdb's ”Top 250 Movies” [1]. The generated results generated are based on the measurement of recommendation percentages for each film.

In the same context, the authors of (Mocanu et al., 2016) proposed a method for the search of filmic documents based on the extraction of subjects. This method is based on the identification of different subjects existing in the subtitles of the filmic documents. In order to identify the topics, the authors considered the following approach. First, a scene temporal segmentation step (Tapu and Zaharia, 2011) associated with the subtitles is set up. Then key words and phrases were proposed to define a term vector for each subtitle. For the definition of this vector, a preprocessing step which considers different strategies of a natural language processing applied to the subtitles, namely, tokenization, deletion of punctuation characters, etc, was applied. Finally, after extracting the key terms, an extraction step of subjects is considered. In fact, the objective of this step is to determine for each term vector a set of subjects. For this respect, a structuring algorithm based on graph is pro-

[1] http ://www.imdb.com/chart/top

posed. Then, the evaluation of this method is tested on a basis of 10 videos selected from a TV archive in France. Then, a comparative study between the results obtained from the proposed method and those obtained through five human observers showed the performance of this method. In fact, the results generated have an average accuracy rate greater than 0.5.

The authors of (Kurzhals et al., 2016) proposed a multimodal method for the analysis of movies in thematic scenes. Indeed, they opted for a low level analysis. Since they proposed a method based on textual documents related to the film such as scripts and subtitles. In fact, the basic objective of this thematic analysis is to explore filmic documents based on the following four aspects: who?, What? and where?. In this respect, the authors suggested combinig the information extracted from textual and visual modalities. Indeed, from the audiovisual stream, they have discussed the detection of plans and movements that capture the scenes. In addition, they selected a set of linguistic structural elements from the scenes of movie scripts and associated them with the subtitles. Based on the various extracted knowledge and interrogation techniques, a semantic analysis technique is discussed. This technique facilitates the comparison between the different scenes on an image and on a semantic shot. In fact, the evaluation of this method is carried out based on a case study of a popular film, namely, ”The Lord of the Rings”.

Then, the study of the methods proposed for the description of the topic of the filmic documents showed their reliability at the level of the used information sources and the considered modalities. Nevertheless, the extraction techniques used here are based on statistical analyzes that cannot semantically provide important knowledge. In addition, a complete evaluation through known film bases is missing. In what follows, we will present the various studies related to the extraction of topics from textual documents.

Actually, the thematic segmentation of textual documents has attracted great attention in the Natural Language Processing domain through several fields, such as, Key phrase extraction via topic, summarization and other disciplines which are interested in the information extraction of textual content. Consequently, several research studies proposed to extract topics or theme documents. These existing methods can be divided into three main classes: Linguistic analysis-based approach, Statistical analysis-based approach and Semantic analysis-based approach.

The first is based on the extraction of linguistic terms or markers in order to detect the definition of a new subject in the text (Rahangdale and Agrawal, 2014). Ragarding the second approach, it was considered very important by some literature studies as it proved to be efficient in detecting the textual topic and thematic. In fact, among the various proposed models we can cite the TextTiling model (Chabi et al., 2011), LSA (Latent Semantic Analysis) model (Bellegarda, 1997) (Atkinson et al., 2014), PLSA (probabilistic latent semantic analysis) (Hofmann, 1999) and LDA (Blei et al., 2003). As it is indicated, the presented models are not really recent but are still very used in the literature, such as, the LDA model.

However, a literature study revealed that each model has constraints. For instance, the latent semantic analysis (LSA) techniqueis effective only for information retrieval tasks, while the probabilistic latent semantic analysis (PLSA) is not suitable for the treatment of documents that do not belong to the training data, On the other hand, the linear discriminant analysis (LDA) uses a very large Bag of Words (BoW) model.

In fact, several research studies proved the efficiency of the LDA model in extraxting hidden topics. For this reason, we propose the use of this model trying to reduce the Bag of Words (BoW) model.

The Outcome: In this literature review, we have presented some research studies that focus on the description of themes and topics based on film and textual documents. More particularly, an examination of the textual documents revealed that the statistical model ”LDA” is the most adapted, the most relevant and the most used for the extraction of topics and themes through textual documents(Jelodar et al., 2019) (Cheng and Hung, 2018).

Despite the shortcomings presented by some studies, we propose to adapt this model in order there are still to improve the performance of the obtained results as well as to overcome the badly treated problems. In fact, our study of the filmic document revealed that several recent studies have been interested in this problem as they have focused on the use of different modalities for the extraction of subjects. Moreover, despite the importance of these studies some shortcomings some of which are caused by the non-use of use of semantic techniques. In fact, since the key terms extracted through low-level characteristics can not present the semantic reasoning of the content, we then opted for a description method based on the semantic analysis. Indeed, the model presented handles the document as a bag of words which allows obtaining a large number of sample data. Accordingly, LDA does not explicitly model the relationship between topics. In this work, we focus on topic extraction by the correspondance process between a list of topic and an ontology to overcome these problems handled poorly.

3 PROPOSED APPROACH

Due to the persistence of the dissatisfaction of the users' needs of the audiovisual documents and the lack of reliability of the studies conducted on this topic, an approach of knowledge extraction is proposed in order to describe audiovisual documents. Such a semantic and automatic description guarantees the interrogation process of reaching a high-level structure. We are interested in the description of filmic audiovisual documents. The input of our approach are a movie script and a set of superposed text on the image.The output are a set of filmic segments. Each sement is defined by a set of knowledges. These knowledges are represented by semiotics descriptions (Dominant theme, taxeme, Discours themes and specified themes) as is indiquated in Figure 8. Therefore, we devote this section to provide an overview of our proposed approach which the contribution finds its originality in:

- The extraction of knowledges from the content. These knowledges are represented as semiotic descriptions of different analytical modalities (textual and visual).
- The automation of the description process.
- The fusion between the modalities to improve the semantic description of the film.

The objectif behind the extraction of this type of kgnowledge is is undoubtly a better interrogation of audiovisual document. In fact, the general principle of our approach consists of three phases: i) the analysis and textual segmentation phase ii) the knowledge extraction phase and iii) the combination and identification of description phase.

3.1 The Analysis and Textual Segmentation

Knowing that the movie audiovisual document involves mainly three necessary phases; a pre-production phase, a production phase and a post-production phase, we find it essential to exploit some documentations used in these phases in order to extract the semantic description. In this paper, we should be based on the documentation prepared in the pre-production phase such as the script of the film through a textual analysis. As this first analysis can hide some realities linked to the content of the film after production, we suggest analyzing the movie in its post-production phase, using the visual modality. A primordial phase in such a process consists of the segmentation and the analysis of the used textual documents, namely, the scenario and the texts superposed on the images.

On the other hand, given that the text in the audiovisual documents represents an important source of description, we interest in (Fourati et al., 2015b) to extract superposed texts from the images of audiovisual documents through the visual modality. The proposed approach comprises of two phases: a learning phase based on the neural network learning and text detection phase. The latter consists of three steps: a pre-treatment step consists in grouping adjacent regions with the morphological operations, then the second step consists in extracting the candidate text areas and finally, the refinement step through a Sobel filter. In the text detection phase, the steps previously cited, are essentially preceded by a temporal segmentation step to obtain the text localization. Once the text is localized, we used the Optical Character Recognition Tesseract (OCR) for the recognition of the superposed text on the images. More precisely, we adopted the impute of Tesseract (OCR) by our result of the process text localization.

Besides, our aim is to take advantage of the discourse and the different knowledge included in the script in order to extract the descriptions. Then, once the inputs of our approach are obtained, we propose an analysis of the discourse present in the pre-production (the script) and the post-production (the superposed text in the image) phases in order to extract the semiotic description proposed by Peter Stockinger (Stockinger, 2013).

In order to segment the script, we are based on the structure of the scripts writed written by the scriptwriter of films. This structure follows a well-defined model that presents each scene separately through a "slug line". In fact, this line starts by necessarily with ".EXT", ".INT", or "scene1", "scene2", etc. The following figure shows an example of a secene with a slug line that starts with "INT.".Based on these lines, we propose segmenting the script into speech units.

Once the speech units are obtained, we will categorize the movie scene in order to bring back the textual segment of the superposed text segment. In fact, we propose to measure the similarity between the superposed texts of the images and the speech units. Therefore the results achieved in this phase are:

- (A): A set of speech units that containt a discourse in each scene.
- (B): A set of superposed texts in each movie segment.

These results were taken as starting points for our

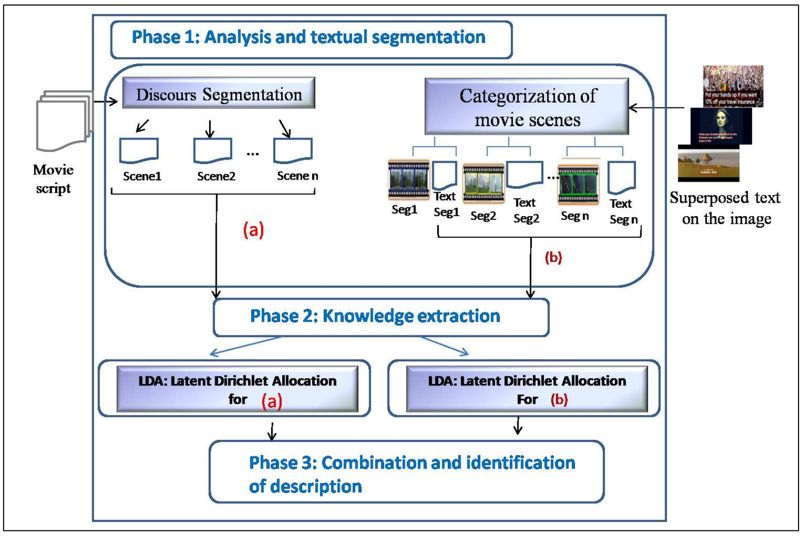


Figure 1: The proposed approach.

INT. DERELICT CASINO - NIGHT

Billyboy gang on stage tearing clothes off a screaming Girl.

ALEX (V.O.)
It was around by the derelict casino that we came across Billyboy and his four droogs. They were getting ready to perform a little of the old in-out, in-out on a weepy young devotchka they had there.

Figure 2: Example of a scene with a slug line.

knowledge extraction phase, in which the semiotic descriptions are extracted from the content of the movie.

3.2 The Knowledge Extraction

Based on the results achieved in the previous phases (a) and (b), we propose in this phase to extract knowledge separately from each result. Therefore, we find it essential to identify both the theme and the topic. Consequently, it is interesting to collect the obtained ideas and knowledge from the state of the art to the extraction of topic and themes. In fact, different related studies proved the performance of LDA probabilistic technique for the extraction of topics from textual documents. The principle of this probabilistic technique is described in the two following steps:

1. Step1: **A Preatreatment Step:** In order to reduce the number of the sample data, we proposed to refine the given input text through the following pretreatment step:
   - The Removal of stop words :all insignificant terms.
   - The Lemmatization of all the terms through the Stanford lemmatizer.
   - The measurement of the pertinence of each term: We propose to measure the relevance of each term in the appropriate segment. We use the metric TF-IDF (Term Frequency-Inverse Document Frequency) as in the previous result proposed by (Fourati et al., 2014) (in the genre and key word extracted from the synopses). In order to extend the scope of the terms, we measure the pertinence of the term with all its synonyms and hyponyms of all genres, which are extracted from the content defined by the Wordnet. In fact, the following algorithm presents the steps to follow when measuring the pertinence of the terms.
2. Step2: **Elicitation of Topics:** The LDA (Blei et al., 2003) represents a probabilistic technique that enables to explore the hidden topic in the textual document with word distribution associated with each topic and its pertinence value. In order to apply this technique, a set of estimated parameters can be manually defined, such as, K, α and β.

   Hence, the general principle of the LDA model is described as follows:
   - K: Represents the number of topics. To select this number, we start with the assumption that a movie segment cannot present more than a topic. In fact, we define $k = n/2$ with n represents the number of segment in the document.
   - α: Represents the priori weight of topic k in a document.
   - β: Represents the priori weight of word w in a topic.

In this phase, to apply the LDA method in each result of the first phase separeatly (LDA with (a) and

Algorithm 1: Pertinence Measure.

```
Input  : XML_File: Represent the keywords
         and the genre,
         Keyword[l]:table of keyword,
         Matr_mot[n,m]: matrix represent all
segment and their term
         Matr_genre[gn,sh]: matrix
represent genre of document and synonym
and hypernyms
Output: Tabword =
         word1_pert, word2_pert, etc,
         Pert[m] : table of pertinence,
         n : number of segments,
m : number of terms et
         l : number of keywords

begin
    k < -0
    M < -Wordnetresult
    for ( i ∈ [1..n] ) do
        for ( j ∈ [1..m] ) do
            if exist(Matr_mot[i], Keyword) =
             true then
                Pert[k] := TF −
                 IDF(Matr_mot[j], Doc)
                 K := k + 1
            else if
             exist(Matr_mot[i], Matr_genre) =
             true then
                Pert[k] := TF −
                 IDF(Matr_mot[j], Doc)
                 K := k + 1
```

LDA with (b)). The result of this step helped obtain a set of terms each of which represents a topic related to the film document. Figure 3 shows a sample of the result of the "LDA" model.

```
topic0 :
night 2.0768431983385254E-4
close 2.0768431983385254E-4
capable 2.0768431983385254E-4
small 2.0768431983385254E-4
call 2.0768431983385254E-4
place 2.0768431983385254E-4
tray 2.0768431983385254E-4
join 2.0768431983385254E-4

topic1 :
feeling 0.016683831101956745
way 0.012564366632337795
call 0.0022657054582904223
like 0.0022657054582904223
home 0.0022657054582904223
night 2.0597322348094748E-4
close 2.0597322348094748E-4
```

Figure 3: Extract of the result of the LDA model.

As mentioned in Figure 3, the result of the adapted "LDA" model shows discrete subjects including a set of keywords. This set of keywords is considered as a set of "specified themes" each of which is segmented to obtain a text from the audiovisual document. Figure 4 shows an example of the result achieved in the previous phase.

3.3 The Combination and Identification the Descriptions

Once knowledges are extracted from each ressource, we focus in this phase not only on the combination between the results obtained from each modality, but also on the identification of the semiotic descriptions. In fact, the herein phase includes essentially two processes: the correspondance between the topic and the theme and the modality combination and identification of the semiotic descriptions.

3.3.1 Correspondance between the Topic and the Theme

As previously mentionned, the obtained result with the LDA technique is a set of abstract topics without the indication of the semantic of each topic. In fact, we chose to use a second technique for the identification of topics in each theme set and keywords. In order to extract the possible semiotic descriptions, we propose to correspond each considered topic to a generic concept that refers to a topic of the movie. In fact, we are interested in the extraction of the concepts of the topic based on a correspondance process between the result of the LDA of each imput and a source of the semantic analysis. In this case, we need a vocabulary of concepts in order to identify the generic concept of a set of themes. Moreover, in this respect, we propose the use of two sources of knowledge as mentionned in Figure 8.

- (A): a theme list of proprietary and non-exhaustive film themes collected from the Internet (1000 themes).
- (B): The ontology LSCOM [2] (Large Scale Concept ontologyMultimedia) (1000 concepts where each concept is described by a definition). Besides, the concepts are organized into six categories: objects, event activities, places, people, graphics, and program categories.

It is on the basis of these sources of knowledge (A) and (B)) as well as on our results of the "LDA" model ((C)) that we propose a solution for the definition

[2] http ://www.ee.columbia.edu/ln/dvmm/lscom/

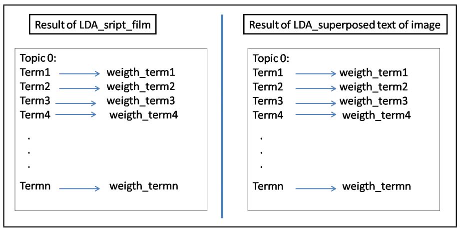


Figure 4: Example of the result of the knowledge extraction phase.

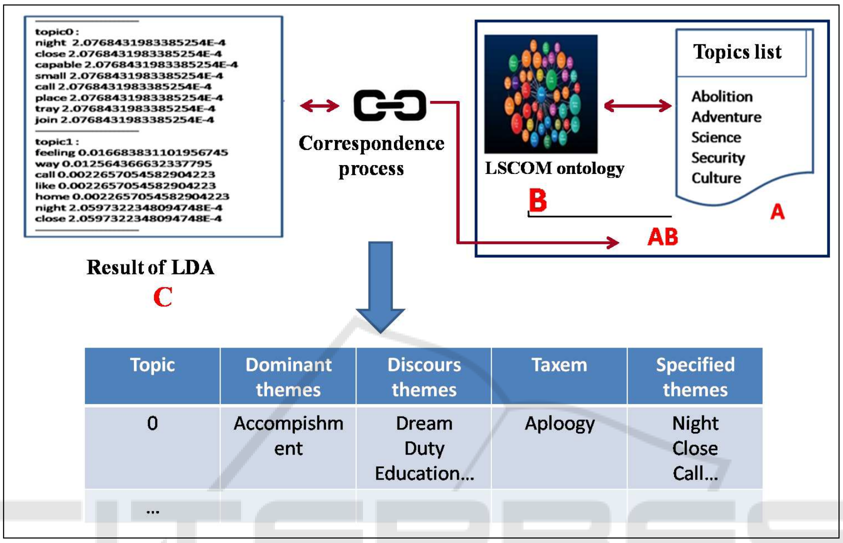


Figure 5: Correspondance process.

of the topic and the different semiotics descriptions. Such a solution is based on two levels of correspondence.

The first level is to match the list of (A) topics with the ontology concepts and their definitions (B). In fact, the main objective of this level is to enrich the semantics related to each theme and etend the thematic concept field.

The second level consists in matching the first level (AB) result with the ones of the LDA model (C). We propose to measure the similarity between themes / concepts in order to discriminate the semantics between themes / concepts. In this respect, several works of literature have demonstrated the importance of the taxonomic measurement "Wu and Palmer" (Dascalu et al., 2013), (Harispe et al., 2014) (He et al., 2016) to measure the similarity between terms and concepts. Therefore, the most similar theme to our results of the LDA model (C) represents the dominant theme. In addition, each extracted theme having a similarity between the result of 'C' and that of (AB) different from zero is considered as a theme of the discourse.

According to Stockinger (Stockinger, 2003), the taxeme represents the relevant theme of the discourse in the topic. To this end, referring to the dominant theme, we propose to measure the "Wu-Palmer's" similarity between the dominant theme and the different themes of discors. As a result, the themes of the discourse which have the maximum value representing the taxeme.

After identifying the concepts related to each set of themes; two cases are then possible:

- If the set of themes is represented by a single concept 'ci', we consider 'ci' as a relevant concept, in other words as a relevant topic for the film. We then define 'C' as a set of the different relevant concepts with C = 'ci'.
- If the list of themes is represented by different concepts, we notice a disambiguation process. Our process is based on the following assumption "an audiovisual segment can not cover more than one subject." To derive the relevant concept from each segment, we apply the following rule:" if a set of themes is represented by different concepts, we measure the similarity between these concepts and all of the relevant concepts 'C'.

In order to measure the similarity between concepts, we use the measure " Wu and Palmer " (Chabi et al., 2011). The process of the thematic description allows the extraction of semiotic descriptions in a semantic way from the content of the filmic documents. Such a semantic description based on the information extracted from the content, becomes a necessary condition for a better interrogation. Figure 8 shows an example of the result of the correspondance process in order to extract the following semiotic descriptions: Dominant theme, themes of discourse, taxeme and specified themes.

This correspondance process is applied firstly with the result of LDA for the script of the film and secondly with the result of LDA for the superposed text of the image. Futhemore, based on these obtained results, we move now to identify the appropriate generic semiotic description for the movie document.

3.3.2 Modalities Combination and Identification of Semiotic Descriptions

The results of the previous process (the relevance measure of each detected topic and the correspondance between topic and theme) should be considered as a starting point for combining the used modalities and in order to identify the semiotic descriptions related to the movie. Figure 6 shows a summary of the result of the previous process.

As a summary, we should recall that we obtained from each source a set of topics each of which is represented by:

- A dominant_theme: Dominant_theme i
- A set of discourse_themes: discourse_theme0, discors_theme1,...
- A taxeme
- A set of specified_themes: sp0,sp1,...

On the other hand, the goal behind modality combination and the obtained results is undoubtly a better description extracted from the movie content. Such a phase is composed of the following two steps: (1) identification of semiotic descriptions and (2) segment structuring.

1. **Identification of Semiotic Descriptions:** In order to identify the semiotic description, two cases are therefore possible:
   - **Case1:** If dominant_theme i of topic i from the script and the superposed text of the image are identical then segment i is represented by:

     seg_i={Dominant_theme_i, {discors_theme}={discors_theme from the script ∪ discors_theme from the script}, {specified_theme}={ { specified_theme from the script ∪ specified_theme from the superposed text of the image}, taxeme= the taxeme that represents the maximum value from the "Wu-Palmer" similarity. } }
   - **Case2:** If dominant_theme i of topic i from the script and the superposed text of the image are different, therefore, we propose to measure the pertinence value of the dominant_theme. To measure this pertinence, we the dominant focus on the relevence of each discourse_theme, which is based on the result value from the "Wu-Palmer" similarity. We consider that:
     Ps i= Wu-Palmer similarity(discourse_theme i, Dominant_theme i of Topic i from the script)
     Pt i= Wu-Palmer similarity(discourse_theme i, Dominant_theme i of Topic i from the superposed text of the image)

     In order to measure the pertinence value of each discourse_theme two cases are then possible:
     - **(i)**: If discourse_theme i exists in topic i from the script and Topic i from the superposed text of the image then :

       $$pertinence\_discors\_theme\ i = max(Pti, Psi) \tag{1}$$

       We consider that discourse_theme i has the highest relevance value as Dominant_theme i. Furthemore, segment i is represented by:

       seg i={Dominant_theme i= discourse_theme that has the highest relevance value, {discourse_theme}={discourse_theme from the script ∩ discourse_theme from the script}, {specified_theme}={ { specified_theme from the script ∩ specified_theme from the superposed text of the image}, taxeme= the discourse_theme that represent the second value from the discourse_theme pertinence . } }
     - **(ii)**: If there is no common discourse_theme in topic i from the script and topic i from the superposed text of the image then we consider that this segment maybe modified in the production phase of the film. The segment is represented by the différent semiotic description extracted from the the superposed text of the image.
2. **Segment Structuring:** The result of the identification of the semiotic description phase is a set of segments "S" with S={Segi} and Segi ={Start, end, Dominant_theme i, discors_theme}, {specified_theme}, taxeme.

   The process used in this phase helps explain the

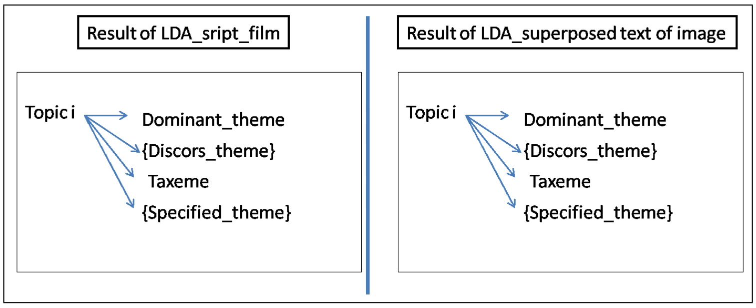


Figure 6: Summary of results.

relationship between different semiotic descriptions. In this phase, our purpose is to structure segment movie documents based on the description that represents a semantic relationship. Such a semantic segmentation , which is related essentially to the description extracted from the content of the document, becomes a necessary condition for linking the document content and description. In fact, we intend to segment and structure movie documents through the topics and the theme. To do this, two types of segmentations, such as, segmentation based topic and segmentation based theme are highlight. The extracted structure is presented in Figure 7 as follows.

4 EXPERIMENTATION

Once the semiotic descriptions of the film content have been automatically obtained, an experimentation process will take place to evaluate the results of the proposed approach. Faced with the absence of similar film methods in the literature to extract the same semiotic descriptions, we proposed in (Fourati et al., 2015a) a semi-automatic semiotic description method to experiment our work. This method is based on the manual annotation.

This section presents the different results of experiments conducted to evaluate our approach of extracting thematic semiotic descriptions. Such a process is based on the extraction of dominant themes, discoures themes, taxeme and specified themes. Two series of experiments are carried out. We therefore concentrate our efforts on the study of the results of the detection of semiotic descriptions through a comaprison of our results with human's judgments. Subsequently, the second series is designed to compare our automatic method to the semi-automatic one. We use the GroupLens[3] database which will be described in the next subsection.

[3]http ://grouplens.org/datasets/

4.1 Data Set

In order to carry out our experimental study, we used the GroupLens dataset. The GroupLens: It is an online movie database which contains basically three sub-databases: a movie database named "MovieLens", a user database and a rating database. In our work, we are interested in the sub-database "MovieLens". Indeed, the latter consists of 10731 films that consider 18 film genres.

4.2 Validation Technique

To evaluate the performance of our solution for the different experiments, we use the following performance measures: Recall, Precision and F-measure. The adapted formulas are:

$$Recall = \frac{Tp}{Tp+Fn} \tag{2}$$

$$Precision = \frac{Tp}{Tp+Fn} \tag{3}$$

$$F\_measure = 2 * \frac{Recall * Precision}{Recall + precision} \tag{4}$$

With:

- Tp represent the number Of similar descriptions identified from our result and the expert result.
- Fp represent the number Of descriptions not identified from our result and identified from the expert.
- Fn represent the number Of descriptions not identified from our result and not identified by the expert.

Then: Tp+Fn represents the total number of pertinent descriptions and Tp+Fp represents the total number of detected descriptions.

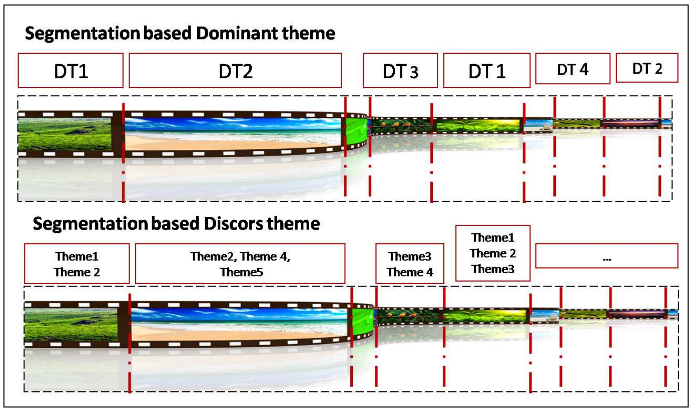
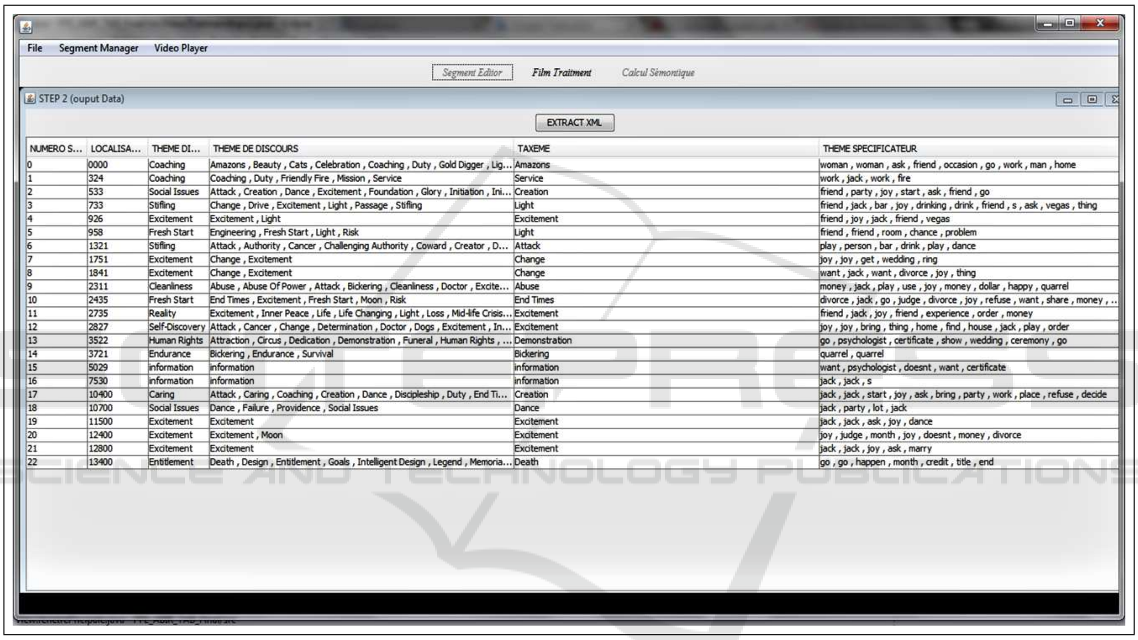


Figure 7: Movie Structuring.



NUMERO S...	LOCALISA...	THEME DI...	THEME DE DISCOURS	TAXEME	THEME SPECIFICATEUR
0	0000	Coaching	Amazons , Beauty , Cats , Celebration , Coaching , Duty , Gold Digger , Lig...	Amazons	woman , woman , ask , friend , occasion , go , work , man , home
1	324	Coaching	Coaching , Duty , Friendly Fire , Mission , Service	Service	work , jack , work , fire
2	533	Social Issues	Attack , Creation , Dance , Excitement , Foundation , Glory , Initiation , Ini...	Creation	friend , party , joy , start , ask , friend , go
3	733	Stifling	Change , Drive , Excitement , Light , Passage , Stifling	Light	friend , jack , bar , joy , drinking , drink , friend , s , ask , vegas , thing
4	926	Excitement	Excitement , Light	Excitement	friend , joy , jack , friend , vegas
5	958	Fresh Start	Engineering , Fresh Start , Light , Risk	Light	friend , friend , room , chance , problem
6	1321	Stifling	Attack , Authority , Cancer , Challenging Authority , Coward , Creator , D...	Attack	play , person , bar , drink , play , dance
7	1751	Excitement	Change , Excitement	Change	joy , joy , get , wedding , ring
8	1841	Excitement	Change , Excitement	Change	want , jack , want , divorce , joy , thing
9	2311	Cleanliness	Abuse , Abuse Of Power , Attack , Bickering , Cleanliness , Doctor , Excite...	Abuse	money , jack , play , use , joy , money , dollar , happy , quarrel
10	2435	Fresh Start	End Times , Excitement , Fresh Start , Moon , Risk	End Times	divorce , jack , go , judge , divorce , joy , refuse , want , share , money , ..
11	2735	Reality	Excitement , Inner Peace , Life , Life Changing , Light , Loss , Mid-life Crisis...	Excitement	friend , jack , joy , friend , experience , order , money
12	2827	Self-Discovery	Attack , Cancer , Change , Determination , Doctor , Dogs , Excitement , In...	Excitement	joy , joy , bring , thing , home , find , house , jack , play , order
13	3522	Human Rights	Attraction , Circus , Dedication , Demonstration , Funeral , Human Rights , ...	Demonstration	go , psychologist , certificate , show , wedding , ceremony , go
14	3721	Endurance	Bickering , Endurance , Survival	Bickering	quarrel , quarrel
15	5029	information	information	information	want , psychologist , doesnt , want , certificate
16	7530	information	information	information	jack , jack , s
17	10400	Caring	Attack , Caring , Coaching , Creation , Dance , Discipleship , Duty , End Ti...	Creation	jack , jack , start , joy , ask , bring , party , work , place , refuse , decide
18	10700	Social Issues	Dance , Failure , Providence , Social Issues	Dance	jack , party , lot , jack
19	11500	Excitement	Excitement	Excitement	jack , jack , ask , joy , dance
20	12400	Excitement	Excitement , Moon	Excitement	joy , judge , month , joy , doesnt , money , divorce
21	12800	Excitement	Excitement	Excitement	jack , jack , joy , ask , marry
22	13400	Entitlement	Death , Design , Entitlement , Goals , Intelligent Design , Legend , Memoria...	Death	go , go , happen , month , credit , title , end

Figure 8: Result system.

4.3 Analysis of the Results

Our focus in this series of experiments on the analysis and the validation of our proposed approach with results defined by an human's judgment. The human's judgments are provided by a domain expert in multimedia analysis. The expert analyzes the audiovisual segments presented in the movie script and based on the list of themes used in our work he gives his opinion. Thereafter, for each film, we compare the discourse theme description extracted from our system with the result proposed by an expert using the same movies indicated in the 'MovieLens' dataset. Table 1 shows an example of the obtained results.

By comparing our system results to the ones defined by a human's judgment, we find that they are satisfactory. In fact, we notice that descriptions are almost similar. Moreover, using the validation technique, we obtained a recall value equal to 88.90%, a precision value equal to 93.04% and an F-measures equal to 90.92 %. These results allow us to conclude that the experimental results are interesting.

4.4 Comparison between our Proposals: Automatic and Semi-automatic Approach

Faced with the absence of similar research studies in the literature, we proposed in (Fourati et al., 2015a) a semi-automatic method based on the manual annotation to extract the same semiotic descriptions ex-

Table 1: Example of the annotation system result.

	Segments	start time	Our System Result	Expert Result
'Day of the dead'	S1	00:00'	y sanctuar	sanctuar
	S2	01:12'	entitlement	action
	S15	100:00:00'	euthanasia	euthanasia
	S19	118:01:00'	gay	gay
	S21	138:00:00'	accomplishment	accomplishment
'The twilght'	S2	06:03	Lycia	Lycia
	S3	09:09'	danger	danger
	S6	15:14'	liberty	liberty
	S14	44:10'	Emotion	Emotion
	S17	117:00'	jazz	information
'The promised'	S3	15:57'	information	recognition
	S4	20:24'	amazons	amazons
	S6	33:52'	duty	duty
	S9	56:41'	lost	lost
	S10	103:00:00'	discovered	discovered

Table 2: Evaluation measure for comparison between an automatic and a semi-automatic approach.

	Automatic	semi-automatic
Recall	88.90 %	90.95%
Precision	93.04%	95.72%
F-measure	90.92 %	93.27%

tracted of our approch. After having evaluated this method by comparing the results with those of a domain expert, we present in this section the result of our experiments. In order to compare the results of the automatic approach with the semi-automatic method, we use the "MovieLens" dataset. Table 2 shows the results obtained by adopting the evaluation techniques more precisely, the : Recall, Precision and F-measures. The results provided for the thematic semiotic descriptions of the filmic documents are satisfactory. Indeed, we managed to segment the audiovisual documents in a thematic way.

5 CONCLUSION AND PERSPECTIVES

As part of the description of the content of audiovisual documents, we proposed an approach that follows a thematic process through different semiotic descriptions.

Recall that among the desirable criteria of a mechanism of describing audiovisual documents of a "film" nature is the integration of one or more modalities. As a result, it is important to extract descriptions from pre-production and post-production filmic documents. Indeed, we proposed in this paper an approach of description of the filmic documents which takes into consideration the various descriptions cited throughout this paper by offering a thematic description approach based essentially on an automatic approach. The input of our approach are a movie script and a set of superposed text on the image.The output are a set of filmic segments. Each sement is defined by a set of knowledges. These knowledges are represented by semiotics descriptions (Dominant theme, taxeme, Discours themes and specified themes) This approach is based on the use of the statistical model "LDA" combined with a semantic technique.

In this paper, two series of experiments are discussed. The first series focuses on the comparison between our system results and those defined by an expert. In the second series, we proposed a comparison between the results of our automatic approach and our previously proposed semi-automatic approach (Fourati et al., 2015a), which enabled us to have more satisfactory results.

Finally, encouraged by our achieved results, we propose in our future research studies to combine the results of the two proposed approaches by combining different sources of information related to the content.

REFERENCES

Atkinson, J., Gonzalez, A., Munoz, M., and Astudillo, H. (2014). Web metadata extraction and semantic indexing for learning objects extraction. *Applied Intelligence*, 41(2):649–664.

Bellegarda, J. R. (1997). A latent semantic analysis framework for large-span language modeling. In *EUROSPEECH*.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Bougiatiotis, K. and Giannakopoulos, T. (2016). Content representation and similarity of movies based on topic extraction from subtitles. In *Proceedings of the 9th Hellenic Conference on Artificial Intelligence*, pages 1–7. ACM.

Chabi, A. H., Kboubi, F., and Ahmed, M. B. (2011). Thematic analysis and visualization of textual corpus. *arXiv preprint arXiv:1112.2071*, 2:1–16.

Cheng, C.-H. and Hung, W.-L. (2018). Tea in benefits of health: A literature analysis using text mining and latent dirichlet allocation. In *Proceedings of the 2nd International Conference on Medical and Health Informatics*, pages 148–155. ACM.

Dascalu, M., Dessus, P., Trausan-Matu, s., Bianco, M., and Nardy, A. (2013). Readerbench, an environment for analyzing text complexity and reading strategies. In *Artificial Intelligence in Education*, pages 379–388. Springer.

Fourati, M., Chaari, A., Jedidi, A., and Gargouri, F. (2015a). A semiotic semi-automatic annotation for movie audiovisual document. In *15th International Conference on Intelligent Systems Design and Applications (ISDA) 2015*, pages 533–539. IEEE.

Fourati, M., Jedidi, A., and Gargouri, F. (2014). Automatic audiovisual documents genre description. In *6th International joint conference on knowledge discovery and informationretrieval (KDIR 2014), Rome, Italy*, pages 21–24.

Fourati, M., Jedidi, A., Hassin, H. B., and Gargouri, F. (2015b). Towards fusion of textual and visual modalities for describing audiovisual documents. *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, 6(2):52–70.

Harispe, S., Senchez, D., Ranwez, S., Janaqi, S., and Montmain, J. (2014). A framework for unifying ontology-based semantic similarity measures: A study in the biomedical domain. *Journal of biomedical informatics*, 48:38–53.

He, Y., Li, Y., Lei, J., and Leung, C. (2016). A framework of query expansion for image retrieval based on knowledge base and concept similarity. *Neurocomputing*, -:Inpress.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57. ACM.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., and Zhao, L. (2019). Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11):15169–15211.

Kurzhals, K., John, M., Heimerl, F., Kuznecov, P., and Weiskopf, D. (2016). Visual movie analytics. *IEEE Transactions on Multimedia*, 18(11):2149–2160.

Martin, J. P. (2005). *Description semiotique de contenus audiovisuels*. PhD thesis, Paris 11.

Mocanu, B., Tapu, R., and Tapu, E. (2016). Video retrieval using relevant topics extraction from movie subtitles. In *12th IEEE International Symposium on Electronics and Telecommunications (ISETC), 2016*, pages 327–330. IEEE.

Rahangdale, A. and Agrawal, A. (2014). Information extraction using discourse analysis from newswires. *International Journal of Information Technology Convergence and Services*, 4(3):21.

Sanchez-Nielsen, E., Chavez-Gutierrez, F., and Lorenzo-Navarro, J. (2019). A semantic parliamentary multimedia approach for retrieval of video clips with content understanding. *Multimedia Systems*, pages 1–18.

Stockinger, P. (2003). *Le document audiovisuel: procedures de description et exploitation*. Hermes.

Stockinger, P. (2011). *Les archives audiovisuelles: description, indexation et publication*. Lavoisier.

Stockinger, P. (2013). *Audiovisual Archives: Digital Text and Discourse Analysis*. John Wiley Sons.

Tang, P., Wang, C., Wang, X., Liu, W., Zeng, W., and Wang, J. (2019). Object detection in videos by high quality object linking. *IEEE transactions on pattern analysis and machine intelligence*.

Tapu, R. and Zaharia, T. (2011). High level video temporal segmentation. In *International Symposium on Visual Computing*, pages 224–235. Springer.